DOCUMENT RESUME

ED 110 491 TM 004 764

AUTHOR Zepeda, Richard A.; Krueck, Thomas G.
TITLE Evaluation of Career Education in Dallas.
INSTITUTION Dallas Independent School District, Tex. Dept. of Research and Evaluation.
PUB DATE [Apr 75]
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS *Career Education; Career Planning; Data Analysis; Data Collection; Elementary Schools; *Evaluation Methods; Information Dissemination; Junior High Schools; Measurement Techniques; *Program Descriptions; Program Development; *Program Effectiveness; *Program Evaluation; School Districts; Senior High Schools; Student Attitudes; Student Evaluation; Testing
IDENTIFIERS *Dallas Independent School District

ABSTRACT

The concept of career education in the Dallas Independent School District maintains that each student should have the option of preparing for a neaningful, fulfilling career while enrolled in school. Career education programs are divided by grade level and stress career awareness (3-7), career exploration (7-9), and career preparation (10-12). Evaluation of these programs provided project personnel with useful information pertinent to operation and the viability of each program as viewed by all segments of the community. Information concerning data collection, analysis, and distribution should provide practitioners with a useful taxonomy of needs in the evaluation of similar career education programs. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED110491

Session 17.03
AERA 1975

EVALUATION OF CAREER EDUCATION

IN DALLAS

by

Richard A. Zepeda
and
Thomas G. Krueck

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Department of Research, Evaluation and Information Systems
Dallas Independent School District
7777 Forney Road
Dallas, Texas 75227

Presented at the American Educational Research Association
Annual Convention
Washington, D.C.
April, 1975

TM004764

Evaluation of Career Education in Dallas

by

Richard A. Zepeda

and

Thomas G. Krueck

Department of Research, Evaluation, and Information Systems
Dallas Independent School District
7777 Forney Road
Dallas, Texas 75227

The Dallas Independent School District made a noteworthy commitment to career education in 1968 with its ground-breaking efforts for the Skyline complex which housed a comprehensive high school, an adult education headquarters, and the Career Development Center (CDC) on an 80-acre campus at a cost in excess of 20 million dollars. Career education plans were expanded and implemented to include all grades, kindergarten through 12, by 1973. By that school year, District monies had been allocated for the trial and evaluation of career education programs at all levels. The evaluation was carried out by the Dallas Independent School District Department of Research, Evaluation, and Information Systems (R&E). The purpose of this paper is to present data which may be generalizable to other evaluation efforts in career education. The first part provides a brief description of each program, including evaluation intent and results. Part two offers a synopsis of the means by which the evaluation was carried out, paying particular heed to methodology data collection, analysis, and reporting. The third part summarizes

---

Presented at the AERA Annual Convention, April, 1975.

conclusions of the evaluation and suggests implications for other evaluation efforts of this nature.

Description of Programs

Elementary Career Education

The elementary career education program (ECEP) in Dallas was a pilot program in its second year of operation. It was implemented in 25 schools in the District and potentially involved approximately 20% of the elementary student population. The program, although admittedly quite small in a district with 138 elementary schools, had been expanded from 87 classrooms in 5 schools to encompass 587 classrooms. The criteria for selecting participant schools for the program were that the principal and at least 50% of the faculty had to express a favorable attitude toward implementation of the program and request to be included in it.

The budget of the ECEP included $4000 in excess of salaries. This small amount maintained resource materials, some of which were available only as examples. Other materials were loaned to teachers by the program and included certain expendable materials, films, filmstrips, posters, teaching units, and kits.

Each cooperating school volunteered a career education representative who was not funded by the program but was a regularly assigned building staff member. Each participating teacher received the program's guidebook which contained, among other things, a synopsis of all career education units which program personnel had

collected from other school districts, and a list of materials suitable for use in implementing career education. The list was organized into three categories: materials available through the program; materials possessed by the program but not in sufficient quantity for dissemination; materials useful but not available through the program, such as particular books or manipulative activities using puppets, hats, and games.

The objectives of the ECEP were several:

1. expand each child's self-awareness
2. develop attitudes about the social and personal significance of work
3. develop and expand occupational awareness
4. improve pupil performance by enlivening basic skills with a career emphasis through "hands-on experiences," exposure to on-site experiences, and classroom visits by people in varied occupations

Evaluation of the elementary program was allocated the resources of one-quarter of a senior-level (doctorate-level) evaluator's time, one-third of an assistant (bachelor-degreed) evaluator's time, and one-tenth of a data clerk's time. The general R&E budget provided typing, printing, keypunching, and computer time.

In keeping with R&E policy for design development, the principal evaluator generated a list of evaluation questions pertinent to the program in August 1973. These questions were to be presented to program management for their consideration and subsequent input. Evaluation designs were due December 1975.

The following method was used to arrive at the evaluation design. In August, an evaluator was assigned the task of evaluating

the program. Since the program had not been the subject of a prior study and since the evaluator was not familiar with it, several meetings with managers were held to inform the evaluator about the program. During these meetings, evaluation questions were agreed upon. Concurrently, a list of information needs of program management was developed. By November the evaluator was familiar enough with the program, and mutual understandings had been reached with program managers so that an evaluation plan could be written. The evaluator and managers cooperated in drafting this plan, a short, written document specifying the questions and information the studies would concern as well as a timetable of events. After this plan had been written, accepted by the parties, and approved by R&E, work was begun on the evaluation design, a longer more detailed document. Program managers and the evaluator cooperated in selecting and designing the instrumentation. The timetable was further delineated. A method of sampling was agreed upon, and an actual sample was taken shortly thereafter. While the writing of the design was done by the evaluator, program personnel did supply inputs and did agree to its substance prior to its formal acceptance as an official R&E design. This document was adhered to as the evaluation continued throughout the school year.

In the ECEP design, process visits were never planned. It was always understood that each classroom portion of the program was unstructured and completely subject to the teacher's volition and discretion. The evaluation unit did not have resources sufficient to monitor 587 classrooms or even a small subset of this number, especially since teachers did not pre-schedule career education

activities. On only one occasion, and then not by plan, did an evaluator see actual classroom utilization of career education resources at this level. Surveys and third parties supplied all other impressions of the program.

Data collected about teachers' opinions were analyzed in a simple manner, using tabulations of percentages and frequency distributions. For the analysis of student data a four-factor post-test only analysis of variance design was used. The dependent variable was performance on the Career Education Self-Awareness Report (CESAR), which is described later in this paper.

After analyzing the data, the evaluator concluded that career education at the elementary level seemed to be a workable educational tool when implemented within the structure of an existing curriculum. Analysis of student data found differences on the attitude test (CESAR) among the variables of treatment-versus-control, sex, ethnicity, and grade. These were statistical differences, and a question existed as to their practical importance. However, the evaluator concluded that future program decisions should take into consideration the grade and sex differences pointed out in the study. A survey of participating teachers found that 75% thought the ECEP to be a worthwhile part of elementary education. Nearly two-thirds of them integrated career education into the regular curriculum rather than teaching separate units. The evaluator reported that teachers accepted and approved of this method. Program managers

learned which components of the program (e.g., materials) were most used and most favored by the teachers and were informed of teachers' opinions about the program.

Junior High Career Education Program

The junior high school career education program (JHCEP) was staffed by a five-person team who worked with the faculty of one junior high school in an attempt to develop a model career guidance instructional system. The system was composed of three sets of objectives and corresponding activities. The primary objective was to develop an instructional model that would integrate career education information into all areas of the curriculum. The model was designed to include performance objectives, learning activities, strategies for implementation, and identification of resource materials. As the secondary objective, the staff was to conduct a year-long series of staff development sessions to encourage and train teachers to incorporate career information into their instruction. Learning of career information by students, the third objective, was to be accomplished by three means:

1. classroom instruction including resource speakers
2. other in-school activities planned by the program developer
3. out-of-school activities planned by the program developers

The JHCEP was funded for an amount of $98,677 including salaries. Evaluation of the program was allotted one-quarter of the time of both a senior and an assistant evaluator and one-tenth the time of a data clerk. Expenses for typing, printing, keypunching and computer time were absorbed by the general R&E departmental budget.

The development of the evaluation design for this program paralleled that described for the ECEP. Both an evaluation plan and evaluation design were written. In general the two evaluations were similar. Having both been done under the branch of R&E known as Developmental Project Evaluation, they followed the same guidelines.

Direct observation of JHCEP operation was confined to sessions with the staff and visits to view materials developed or collected by them. Interviews with the program staff regarding the development and implementation process, along with site visits by evaluators to ascertain the actual availability of supporting instructional materials were the sources of information rather than classroom visits. Although it was anticipated in the program's evaluation design that students as well as teachers would be surveyed and that more detailed observation of the classroom instruction process and staff-development would be undertaken, the problems faced by the program in impacting these areas made these activities unnecessary. In other words, since the JHCEP had such great difficulty in reaching the classroom, evaluation personnel were left with little to observe.

Student data were analyzed statistically. Other data were presented using descriptive statistics. For the statistical analysis of student data, multivariate covariance analysis compared the performance of program and control students over four criterion measures to be described later. All students were equated statistically over the concomitant variables of initial academic ability (operationally defined as Comprehensive Tests of Basic Skills [CTBS] Reading scores), age, sex, and ethnicity.

At the junior high school level the evaluator found there were no differences between students in the program and controls on three of the four criterion variables. In the School Attitude variable a difference existed in favor of control students. The lack of differences in favor of program students on attitude and knowledge measures was ascribed to the failure of the first year of an experimental infusion model. In short, the JHCEP attempted to implement a poorly defined infusion model for which initially there was a minimal curriculum component. Once this trend was reported to them, the staff began modifying the infusion model. Probably the most important result of program activities was to reformulate the infusion model and prepare for the follow-up-phase the second year. The program staff settled upon the eighth-grade curriculum in Social Studies, Mathematics, and Science as a vehicle for the infusion of career education. They were convinced that a program of career information would make senior high school experiences more meaningful and provide for more rational career decisions.

Senior High Level - Skyline Career Development Center

The Dallas high school career education effort is housed in a mammoth facility, Skyline Career Development Center (CDC), opened in the fall of 1971. The facilities and equipment were modern and representative of those found in the various careers for which the students received training. The CDC was conceived as an opportunity for students with appetites for study and learning beyond the conven-

tional curriculum. This opportunity included a wide range of vocational and technical subjects and training in traditional academic subjects so concentrated as to be considered vocational. The Center's distinguishing feature was its provision for extensive preparation in selected career areas without sacrificing college entrance requirements. An acknowledged purpose of the CDC program was to qualify students for employment upon graduation. It was also anticipated that the program would improve student and parental attitudes toward schooling.

Application to the CDC was open to all upper-secondary students in the District. The only stipulations placed upon any applicant were that each have a definite educational goal and the potential ability to accomplish it. For the 1973-74 school year, the CDC had an enrollment of approximately 2500 high school students, representing all 18 District high schools, who studied in 28 clusters which represented a wide variety of career areas. Over 110 instructors were assigned to the Center. Each taught two, three-hour sessions. An extensive administrative staff managed its operations, supervised instruction, and acted as a liaison with Dallas' central administration and the Dallas community, both business and residential.

In order to evaluate such an extensive program, the Board of Education provided a full-time staff of seven persons: two doctorate level evaluators (one functioning as the coordinator), two assistant evaluators, a programmer, a secretary, and a data clerk.

Both the CDC and its evaluation staff had co-existed the previous two years, which allowed for not only greater rapport between project management and the evaluation team, but also an understanding as to what should be considered and which methods would be most appropriate in arriving at a suitable evaluation design. Given this situation, the evaluation team replicated pertinent evaluation questions from the previous year's design and suggested new areas as worthy of examination. Program managers were asked to respond to the merits of all questions. Furthermore, program management identified additional areas of concern as well as those questions of a low priority for which their suggested concerns might be substituted. The final form of the evaluation design was the sole responsibility of the evaluation team with the understanding that the rationale and defense of the questions included were up to the evaluation team. This responsibility was particularly true in any areas where program management and evaluators disagreed. All considerations concerning sampling, analysis, and reporting were the sole responsibility of the evaluation team; however, project management was kept abreast of these decisions.

The major thrust of the evaluation efforts was to (a) create and maintain an extensive and accurate data base on students and instructors, (b) monitor program implementation, (c) assess attitudes and perceptions of various groups pertinent to the CDC and related topics, (d) implement a variety of tests intended to measure

student performance and success, and (e) audit the curriculum development effort.

Evaluation results showed a continuing trend toward greater minority representation in the CDC, particularly by black students. For the third straight year students in general and minority students in particular viewed the CDC in a most positive light. Surveys showed that they perceived job preparation, an opportunity for specialization, individualized instruction, assistance in career selection, and outstanding instructors and facilities as the CDC's most positive aspects.

As might be expected, most students tended to score extremely well on the specific tests within the Armed Services Vocational Aptitude Battery (ASVAB) related to their clusters. Results of attitude measures indicated that enrollment in Career Education provided greater meaning to a student's entire program of studies. Follow-up studies showed that students were able to obtain jobs directly related to their experiences in Career Education programs.

Procedure

Evaluation Designs

For the 1973-74 school year, R&E made a distinction between evaluation plans and evaluation designs. An evaluation plan usually ran to two typewritten pages. It listed a projected number of reports, due dates for them, and the list of recipients. The types of analyses anticipated were noted as well as variables of interest. Data collection methods were also briefly described.

Such an evaluation design took literally months to write. It could run in excess of 50 typewritten pages. Not only were evaluation reports assigned due dates, but also the intent of the reports was outlined with exceeding detail. Each study's sample was described, including a depiction of how the sample was arrived at. The population that the study was supposed to generalize to and any assumptions and limitations of the study were specifically identified. A rather thorough description of the program was provided, a description that could not be gained merely by reading a proposal or some such document. Instead, visits to the program and discussions with its staff, as well as literature reviews, were needed to achieve such a degree of specificity. In general, all details of the evaluation were spelled out. Ultimately, formal designs for the studies discussed in this paper were approved by the R&E chief administrator. In previous years such documents were actually approved by the Board of Education. An evaluator was expected to follow this document to the letter, or document why it was not possible. This year R&E policy has come to rely on only shorter plans, which are not so program specific as designs. It is interesting to speculate upon why so much effort was previously spent to produce evaluation designs. It may have had to do with credibility. Any R&E evaluator could write a plan for any program without knowing very much about it; a feasible evaluation would be described in general terms. The earlier insistence upon a design

guaranteed that an evaluator knew the program both through the local managers and from a knowledge of the literature. At first the Board of Education insured this by a review and approval of the design. Later R&E was trusted to vouch for the integrity of the evaluation by its approval of a very detailed design. Now, apparently, R&E credibility is unquestioned. The Department has proven itself in the sense that evaluation reports are now assumed to be sound studies firmly based upon scientific principles.

The question remains as to whether insistence upon a rigorous evaluation design ultimately results in a better study. Writing the design actually requires almost as much work as writing the final report. We believe that R&E will insist upon the quality of the report; it will not be allowed to suffer. Time formerly expended upon writing the design will thus be spent on writing more interim reports. This policy will also allow for modification and refinement of the methods used in the final data analysis during the school year, without violating self-imposed constraints set forth in a design months earlier when an evaluator was operating under more uncertainty. Although designs are a good test of an evaluator's abilities, they cannot be justified on this basis alone if the overall quality of Departmental studies remains as high as it is now.

Dallas R&E evaluations, whether based on designs or plans, have followed the same general model: Stufflebeam's CIPP model. The research and evaluation unit has tried to operationalize this model whenever feasible. However successful its implementation,

terminology from the model has been employed to categorize areas of evaluation. The areas included in all levels of career evaluation have been context, process, and product evaluation.

The purpose of context evaluation was to describe accurately the setting in which each program was being implemented. This included obtaining and reporting pertinent information relative to the student and instructor populations. Information gathered for the context data base facilitated subsequent evaluation efforts, allowing the findings of product evaluation to be more easily interpreted.

The object of process evaluation was to detect and/or predict program defects during implementation stages. This form of evaluation was possible only after a course of action had been determined and implemented, (i.e., when a program was off the drawing board and in the classroom). When functioning, it provided necessary periodic feedback to project managers and other decision makers who had the responsibility for the program. Process evaluation performed a quality-control function as it monitored a program, paying particular attention to communications, inter-personal relationships among staff and students, and adequacy of the curriculum.

The goal of product evaluation was to measure and interpret student performances as often as possible during the program cycle and at the end of it. Behaviors associated with the objectives of the program were measured and compared with pre-determined criteria. In all programs, a great deal of the information sought for product evaluation relied on psychological testing.

Data Collection

Context evaluation. Having previously noted the purpose of context evaluation, it should be apparent that the initial step toward that purpose was deciding what data should be collected on all populations being considered. While the content of the data bases varied with each program, the following list of student and instructor variables collected for the program at the secondary level is typical.

Student Variables:

- Identification Number
- School
- Date of Birth
- Parent Name
- Address
- Phone Number
- Sex
- Ethnic Background
- Grade Level
- Parent Occupation
- Achievement and Aptitude Test Scores
- Social Security Number

Teacher Variables:

- Identification Code Number
- Sex
- Ethnic Background
- Teaching Experience
- Educational Background
- Non-teaching Experience

Student variables such as socioeconomic status and grade-point average were excluded because of their inaccessibility and questionable reliability or meaningfulness if attained. In these cases, indicators were sought as alternatives. For instance, occupation of the head of household was collected as an alternative to less reliable socioeconomic status data.

The decision as to what variables to include created the more complex choice of how to obtain the required data: either accept previously collected data (noting whatever limitations accompany it), or collect new data. Existing records provided data on such variables as age, ethnicity, sex, date of birth, and standardized test scores. The District's extensive computerized records on each student supplied accurate data on students' demographic variables as well as the District's unique identification numbers, which were cross-referenced with files that stored information about academic performance either over behavioral objectives at Skyline CDC or on standardized test data. The evaluation units thus built data bases that were a subset of the approximately 150,000 records-per-file (one for each student) maintained by the District. Preparatory to analysis, these smaller data bases were maintained, edited, and updated throughout the year.

The collection of data on variables such as head-of-household's occupation required contacting the subjects. As part of this procedure, all other records in the subject's file were validated by the personal inspection of each subject. It was interesting to note that each editing pass inevitably updated program files. Steps were taken to alert management of any changes that had somehow evaded the scrutiny of attendance and counseling offices.

Ultimately, it seemed most advantageous to have the routine District data collection meet the needs of all parties requiring data from students. After one collection, all variables would be stored in a master data file available to each department, office,

or individual. Proposals were submitted to pilot such a system that would eliminate repetitious collection of data. But this system never got past the planning, primarily due to bureaucratic inertia. Each subject still fills out each department's form. At the secondary level a student might fill out as many as six separate cards or forms, each requiring the same variables of interest.

Process evaluation. To document program changes occurring during implementation or to ascertain whether the program was ever implemented, process evaluation centered on monitoring the program. This allowed for flexibility in the program's design which insured that the final product would encompass all stated objectives or modify them in a way dictated by experience.

The thrust of this stage of evaluation included examination of communication channels, interpersonal relationships between staff and students, and adequacy and use of material resources and physical facilities. Process evaluation posed four major questions: (a) to what extent was the program, as designed, implemented? (b) to what degree was the curriculum tailored to individual student needs? (c) did any deficiencies hinder full program implementation? (d) were any innovations taking place that could be generalizable to other classrooms?

Dividing process evaluation into a series of observations in each instructional area of interest seemed most meaningful. Days of the week and times of the day were varied, whenever possible, to reduce the possibility of biasing observations. Observations

generally took from 15 minutes to one hour, depending on the intent of the given observation.

A series of observations noted (a) the types of activities that were taking place, (b) the extent of individualized instruction, (c) the extent of curriculum implementation, (d) any disruptive influences observable in the classroom, (e) the choice of activities for each behavioral objective for the students, (f) the extent of student choice of topics or areas of specialization within the cluster, and (g) any deficiency of supplies preventing students from completing the objectives.

This form of evaluation was very costly with regard to human resources. It was impossible for some program evaluators beyond a token effort. However, some method was necessary to insure that the program was being implemented. Management reacted to process evaluation with cautious acceptance. It was not easy for them to adjust to this scrutinizing examination, yet they too perceived the value of an external, objective observer.

Product evaluation. Year-end reports are expected in Dallas as elsewhere. Program managers use them to assess their degree of success. Most importantly, Board of Education members use them to decide the future of programs, their continuance or expansion.

Product evaluation reports intended to summarize the year's activities for a program and to document each program's successes and difficulties by answering a series of questions. At the beginning of the year the questions had been formulated and agreed upon as

important. In general, evaluation questions asked, "What were participants' (students and teachers) attitudes toward the Program?" and "What effects did the program have upon psychological variables of interest?"

Program effects as indicated by changes in attitude or knowledge were measured in the main by paper and pencil instruments. Attitude instruments were developed within the Department. The R&E staff located, purchased, and sometimes modified standardized tests. Criterion-referenced tests were also located or developed by R&E. Participating teachers or R&E personnel administered these measures, depending upon the circumstances. Scoring was done on a computer using punched card records of item responses preparatory to analysis. Then, needed measures of academic performance, such as CTBS scores, were assimilated from District computerized files.

In addition to data from psychological tests, the programs needed information about the attitudes, impressions, and opinions of the various groups associated with the programs. Survey research approaches to data collection were found to be most suitable for this task.

The most popular method employed was a short self-administered questionnaire/opinionnaire, designed to examine a specific topic. These instruments, developed by evaluation personnel, rarely exceeded 25 questions over one or two pages. It was found that closed-end questions provided higher response rates and more useful information than open-end questions, because many respondents

failed to take the time necessary to compose meaningful answers or because a question's intent was misinterpreted. Multiple choice, agree/disagree, and Likert-type questions were found to be more useful in these programs. It was also noteworthy that rank-ordering techniques were judged unsuccessful because respondents failed to read and follow directions.

More personalized forms of survey research such as interviews and telephone surveys were employed at the Skyline CDC. Report recipients valued information gained from personal contact with students, parents and instructors, and this type of data collection was available to the large Skyline evaluation team. Still, personalized forms of survey research required the reduction of sample size, although sample size was never allowed to become too small for meaningful inferences to be made.

Testing

To gather student data for the final report, the ECEP evaluation used an attitude instrument to measure hypothesized changes in students participating in the program compared to changes in students in the control group. Members of the Research and Evaluation Department developed the Career Education Self-Awareness Report (CESAR), which was designed to measure attitude toward work and self. It consisted of 16 short, direct questions concerning how a student feels when an adult comes to school to be interviewed by students about his or her work, how a student feels about working,

and what a student thinks others feel about him or her. The students were asked to respond to the questions by marking a modified Likert Scale. No cognitive tests were administered at the elementary level.

At the junior high school level, four instruments were used to measure students: (a) the Career Maturity Inventory (CMI) Competence Test (Short Form), which assessed general knowledge of career information, (b) the CMI Attitude Toward Work (Short Form), which assessed student attitude toward work, (c) a School Attitude Scale, and (d) a Self-Concept Scale, which assessed student attitudes toward their own academic performance.

The CMI Competence (Short-Form) was a 20-item multiple-choice test constructed by randomly sampling 20% of the items (i.e., 4) from each of the five subtests of the standardized Career Maturity Inventory published by the California Test Bureau. Only four of the 20 items on each subtest were sampled in order to keep the administration of the test within a 30-40 minute time limit. Although subtest scores alone could not be interpreted, it was assumed the pooled items as a whole provided a valid measure of knowledge of career information and related decision strategies.

The CMI Attitude Toward Work (Short-Form) was a 20-item True-False inventory constructed by randomly sampling 40% of the items from the CMI Attitude Toward Work test, also published by the California Test Bureau. Like the CMI Competence, the CMI Attitude was shortened to fit the allotted two-class-period testing interval. Although the shortened versions made the test norms in-

applicable, the shortened instruments remained highly valid for (experimental) comparison purposes.

The School Attitude and Self-Concept Scales, developed during the school year by R&E personnel, consisted of two sets of 11-item Likert-type statements (to which students indicated agreement or disagreement) arranged in an alternating sequence. The items composing each scale were based upon "item formats," which were in turn derived from a more theoretical analysis of the behavioral referents of "school attitude" and "self (academic) concept" upon which the validity of the scales was founded.

At the CDC, testing involved both norm-referenced and criterion-referenced instruments. The only test administered to all CDC students was the ASVAB, the second year for such testing. The ASVAB combined nine sub-tests into several aptitude areas, since combinations yield a better prediction than individual tests used alone. The Aptitude Areas included Electronic, General Mechanical, Motor Mechanical, Clerical/Administrative, and General Technical. The battery was viewed primarily as a guidance instrument.

Additional norm-referenced testing was conducted in 16 different curriculum areas at the Skyline CDC. A total of 531 students were tested in their particular areas. The tests included the following:

Short Occupational Knowledge Tests
Graves Design Judgment Test
Cooperative English Tests
Emporia Foods Test
MLS Cooperative Foreign Language Tests
Cass-Sanders Psychology Test
Sare-Sanders Sociology Test
SRA Clerical Aptitudes

Some items in the standardized tests were not applicable to the topics covered in the various cluster curricula. Instructors identified such items and they were dropped from the analysis.

Criterion-referenced tests used at the CDC were developed within the program's curriculum efforts. The testing intended not only to ascertain student success but also to act as a pilot study of the tests. Since tests for many of the cluster areas were examined during the 1972-73 school year, only 14 of 28 clusters were represented in the criterion-referenced testing for the academic year of 1973-74. Matrix sampling techniques were used for paper-pencil tests so that each instrument was administered, but every student was not tested on every test each qualified for. Items involving performance of a task were sampled. Attempts to test students on a 25% sample of performance items were generally unsuccessful because instructors did not cooperate.

Criterion-referenced testing presented major difficulties. It had been assumed that this form of testing would provide student success measures, but the tests were generally found to be of questionable worth; so testing was minimized and results were considered meaningless.

Several general comments about testing should be aired. Attitude scales continued to be used although their analyses consistently revealed no difference between the groups studied. Because program personnel continued to write such objectives as attitude changes into proposals, it became almost mandatory to evaluate over these

objectives. Program managers as a class seemed to think such objectives desireable and attainable; and measurement of attitudes does continue to be an area of scientific interest. Developers continue to report new tests in this area. Dallas' evaluators felt that the attitude measures employed were for the most part the equal of what had been developed and reported. However, the finding of no difference remained almost a foregone conclusion.

A second general comment, cognitive tests that related directly to a program were often impossible to locate. This led to either the development of such measures or the use of proxy instrumentation. The development of such instruments at times meshed quite nicely with other efforts of the Department since R&E maintained a criterion-referenced test development unit, which developed tests specific to the Dallas curriculum. But sometimes the evaluator became a test developer out of necessity. Tests developed and used during the period of an evaluation study, less than a year, necessarily had shortcomings.

A third and last general comment, even when tests were procured or newly developed, instructors and managers at times resisted their use, often basing their resistance on arguments about impingements upon instructional time. Other disclaimers focused on the evaluator's lack of background in a particular subject and thus impugned his expertise in test development in that area.

Analysis

The form of the data dictated the analysis used. Efforts

were made to conduct analyses so that readers would find the information most useful. Types of data and the subsequent analyses used are identified below. They range from the most descriptive forms, involving means and standard deviations, to the inferential forms of analysis, including multivariate techniques.

With information obtained via survey research, the evaluators generally chose to report proportions and frequency distributions with an occasional Chi-square test for statistical significance. It is well known that statistical differences are affected by sample sizes; therefore, observable differences often had greater meaning to the readers, even though the evaluators cautioned that observable differences should be viewed carefully since they might be due to chance. Observational data received no formal analysis. It was used to present program management with a summary report of the data in which impressions were synthesized and transcribed into a narrative.

In general, where a "higher" level of data was collected, a more sophisticated level of analysis was used. For the analysis of both attitude instruments and knowledge instruments the evaluators followed classical, statistical methodology. Within the constraints of the program, random samples, sometimes stratified according to variables of interest, were drawn. Instruments were then administered according to a predetermined plan with serious attempts made to standardize test administration.

Where one dependent variable was of interest, a general linear model was used with categorical variables such as sex, ethnicity,

and grade-placement included in the model. Statistical tests of hypothesis about the variable were then conducted. Typical ANOVA tables were presented to the readers. Inferences about the results of statistical tests were reported.

When multiple criteria were analyzed, a multivariate covariance analysis was used to test hypotheses. The multivariate model for JHCEP included ethnicity, age, and sex as well as achievement scores. Significance tests were conducted between the experimental and control students on the dependent variables of career knowledge, work attitude, school attitude, and self-concept. Tables of means and standard deviations for the subsets by groups were presented and the results of hypothesis testing explained.

In one case where attitude instruments were developed by the evaluator, an analysis of the instruments using data collected prior to the evaluation study was done and described. The instruments were analyzed by principle components factor analysis followed by a rotation. The factor structure resulting from this analysis was then illustrated. Reliability and validity coefficients were calculated and presented.

It could be argued that the more sophisticated the analysis the more it was done for the evaluator and the less it was done for the layman reader. There was little doubt of the R&E staff's ability to write at a level which would be of limited or no use to most District decision makers. It was never an avowed intention

to overwrite the audience. The opposite was true. However, because of training, the inclination toward esoterica on the part of evaluators remained strong. Probably most R&E studies could have been accomplished with nothing more than t-tests, which most of the audience could readily understand. Yet it often seemed that highly trained persons exercised their skills even at the expense of doing more work--conducting a more difficult analysis than necessary and then writing extensively to explain it.

Reporting

Two most important considerations for evaluators were the method of reporting and who should receive the reports. R&E policy is to forward all reports to each decision maker who needs and uses them. Consequently, reports were written so that the recipients would understand and be able to deal with the findings presented. This meant that more esoteric approaches to analysis and reporting were sometimes bypassed. Reports were distributed to various levels of administration so that the information disseminated through the organization. They were released to individuals at the Assistant Superintendent's level, Deputy Assistant Superintendent's level, and Program Manager's level. All final reports were channeled through a team of top-level District administrators to an evaluation committee of the Board of Education. Ultimately, they were presented at a public, general meeting of the Board of Education.

Recipients handled reports critical of the program quite differently from reports noting favorable aspects. This was particularly

true with process evaluation reports which, by the nature of such evaluation, created difficulties for program managers who were highly accountable to their superiors in the administrative hierarchy. Occasions arose when memoranda had to be issued on how certain reports were being misused to the extent of threatening lines of communication between evaluators and program personnel.

It should also be noted that verbal forms of communication became important. Information not conforming to evaluation guidelines but deemed so urgently necessary for program considerations that it could not wait for normal written reporting, or information of a nature that precluded reporting in print was released via informal verbal communication.

Conclusions and Implications

Evaluation of Career Education in Dallas consisted of studies of three programs - ECEP, JHCEP, and CDC. In the ECEP the evaluator found differences between the treatment and control groups on the attitude test, CESAR. He concluded that future program decisions should consider the documented grade and sex differences. A survey of ECEP teachers found that 75% felt the program was a worthwhile part of elementary education. In the JHCEP the evaluator determined that the program was scarcely implemented. The evaluator concluded that the lack of differences in favor of program students was due to the failure, during the year, of an experimental infusion model. The most gratifying part of this evaluation was that once the failure was pointed out the staff began modifying the model. By the end

of the school year, the infusion model had been reformulated for another trial the following year. At the CDC students in general, and minority students in particular, viewed their experiences there in a most positive light. The evaluator concluded that enrollment in career education added greater meaning to a student's entire program of study. Surveys showed that students perceived job preparation, the opportunity for specialization, individualized instruction, career guidance and outstanding instructors and facilities as the most positive aspects of the CDC.

In the course of these evaluation studies, the authors learned, or relearned, much about evaluating career education programs. Some observations are presented here in the hope that they might be useful to others charged with program evaluation.

Regarding the development of an evaluation design, the authors conclude that while the task of developing a detailed evaluation design is a good measure of an evaluator's writing skills, it certianly does not guarantee a superior final report.

Remembering that the planning of a scientific study involves the problem of data collection, the authors support the concept of a centralized student data bank which minimized duplication of effort in data gathering. Realizing, however, that this is the ideal and not the usual situation, the authors advise careful consideration of these problems (a) the variables to be considered in the analysis, (b) the collection methods to be used, (c) the task of building a data base from disparate sources, and (d) the responsibility of editing the data.

It is certainly worthwhile to make a conscious effort to establish and maintain close communication between program managements and evaluation staff. In the programs discussed, meetings were held during which evaluators became familiar with the problems and information needs of program administrators. Keeping in mind these needs, the evaluator devised the formal evaluation scheme. As a result, management was predisposed to the credibility of the evaluation reports.

When considering the continuation, modification, or expansion of programs, the Board of Education utilized product evaluation information. Such information, they felt, allowed them to assess the success of the programs under study. Usually the information was presented as measured behaviors, behaviors associated with program objectives, which were compared either to pre-determined criteria or to a control group.

Invariably much of the product data was the result of psychological testing with which were associated certain problems. Evaluators frequently found it impossible to locate cognitive tests which related directly to a program. Consequently the evaluator necessarily became a test developer. Tests developed by R&E personnel and used during the period of an evaluation study, less than a year, necessarily had shortcomings. Tests developed by program staff also often presented problems. This was true especially at the CDC where it had been assumed that criterion-referenced tests would provide student success measures. When

these instruments were found to be of questionable worth, testing was minimized, and results were discounted.

Because of program developers insisting that their programs would affect student attitudes it was required that attitudinal measures be utilized in the evaluations. The authors believe that, for the most part, the measures employed were equal to those developed elsewhere. However, a finding of "no significant difference" was almost a certainty in any program R&E evaluated.

Report recipients seemed most to value information gained from personal contact with students, parents and instructors. Unfortunately the human resources necessary for collecting this type of data were available only at the CDC. Additionally, personalized survey research required that the evaluator reduce the sample size with the resulting trade-off of statistical power.

It is the authors' opinion, because of possible benefits to the program, that a large allocation of human resources to process evaluation is the most advantageous alternative. Process evaluation documents changes occurring during implementation, confirms that implementation actually occurred, and enables managers to make meaningful decisions to improve the program during the year.

Program managers were found to be most interested in immediate kinds of information (i.e., context and process reports). Usually they wanted information on what was currently happening in their programs. The evaluator's concern was to get this information to them as soon as possible after it was gathered. Although they were interested in all types of information, program personnel, it was

learned, were most interested in what happened yesterday in the classroom or at teachers' meetings than in year-end measures of student cognition.

On the other hand, elaborate statistical designs allow the evaluator to conclude that students enrolled in the program were or were not affected by some treatment. But such studies tend to have little impact on a program unless dramatic differences are found. Even when such differences occur, almost certainly the students did not receive the treatment as it was originally proposed. So the authors recommend a reduced emphasis on elaborate statistical designs when trade-offs have to be made as they certainly will.

It is likely that most research and evaluation studies, including those discussed in this paper, could be accomplished by using easily understood t-tests. However, evaluators, trained to display their mastery of statistics and experimental design, chose sophisticated designs for evaluations and utilized esoteric analyses. They then must write extended explication for the benefit of those who are to utilize the findings.

The reporting of information also has associated problems. Sometimes findings which were considered urgent were communicated verbally. At other times, information which did not fall within the specified scope of the evaluation or information otherwise considered inappropriate for normal reporting was also communicated verbally. Of course the difficulty of making reports timely was ever present and on occasion memoranda were issued to discuss how certain reports were being misused.

It is the authors' wish that some of the candid comments and incidents included in this paper will suggest to the reader what is career education evaluation in Dallas. Hopefully, sharing mistakes made and lessons learned in Dallas can contribute to more meaningful evaluation throughout the educational community.

REFERENCES

Dagley, P. R. and Zepeda, R. A. Final evaluation report for career education in the elementary schools. Research Report No. 74-281, July, 1974, Department of Research, Evaluation, and Information Systems, Dallas Independent School District.

Dallas Independent School District; Department of Research, Evaluation, and Information Systems. Abstracts of research and evaluation reports 1973-74. Dallas: DISD, 1974.

Krueck, T. G. Plan for the evaluation of the Skyline Career Development Center 1973-74. Research Report No. 73-211, October, 1973, Department of Research, Evaluation, and Information Systems, Dallas Independent School District.

Krueck, T. G. Final product evaluation report for the Career Development Center, 1974. Research Report No. 74-262, July, 1974, Department of Research, Evaluation, and Information Systems, Dallas Independent School District.

Vitale, M. R. Evaluation design for the Title III Dallas jr. high school career education project 1973-74. Research Report No. 74-222, December, 1973, Department of Research, Evaluation, and Information Systems, Dallas Independent School District.

Vitale, M. R. Evaluation report for the Title III Dallas jr. high school career education project 1973-74. Research Report No. 74-291, July, 1974, Department of Research, Evaluation, and Information Systems, Dallas Independent School District.

Zepeda, R. A. Evaluation design for career education in the elementary school. Research Report No. 74-217, December, 1973, Department of Research, Evaluation, and Information Systems, Dallas Independent School District.